

WILLCOME INSTITUTE LIBRARY	
I.	welMOrnec
I	pam
	WA 400
	1924
	Y 95f

Report No. 28

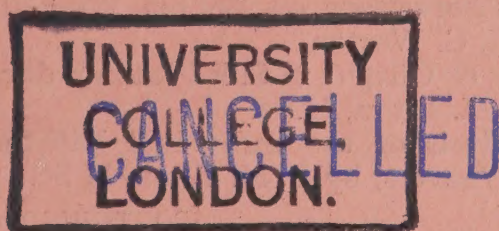


MEDICAL RESEARCH COUNCIL.

INDUSTRIAL
FATIGUE RESEARCH BOARD.

The Function of Statistical Method in Scientific Investigation.

By G. UDNY YULE, C.B.E., M.A., F.R.S.



LONDON :

Published by His Majesty's Stationery Office, and to be purchased
at any of the addresses shown overleaf.

1924.

Price 6d. Net.

CANCELLED

A 306 MED

OTHER REPORTS OF THE INDUSTRIAL FATIGUE RESEARCH
BOARD.

A.—General and Miscellaneous.

- First Annual Report to March 31st, 1920. *Price 6d. net.*
Second Annual Report to September 30th, 1921 (with Analysis
of Published Work). *Price 1s. 6d. net.*
Third Annual Report to September 30th, 1922. *Price 2s. net.*
Fourth Annual Report to December 31st, 1923. *Price 1s. 3d. net.*
No. 4.—The Incidence of Industrial Accidents, with special reference to
Multiple Accidents, by Major M. Greenwood and Hilda M.
Woods. *Price 6d. net.*
No. 6.—The Speed of Adaptation of Output to altered Hours of Work,
by H. M. Vernon, M.D. *Price 1s. net.*
No. 12.—Vocational Guidance (A Review of the Literature), by
B. Muscio, M.A. *Price 1s. net.*
No. 13.—A Statistical Study of Labour Turnover in Munition and other
Factories, by G. M. Broughton, M.A., E. H. Newbold, B.A.,
and E. C. Allen. *Price 3s. net.*
No. 14.—Time and Motion Study, by E. Farmer, M.A. *Price 2s. net.*
No. 16.—Three Studies in Vocational Selection, by E. Farmer, M.A.,
and B. Muscio, M.A. *Price 1s. 6d. net.*
No. 19.—Contributions to the Study of Accident Causation, by Ethel
E. Osborne, M.Sc., H. M. Vernon, M.D., and B. Muscio, M.A.
Price 1s. 6d. net.
No. 25.—Two Contributions to the Study of Rest-pauses in Industry, by
H. M. Vernon, M.D., and S. Wyatt, M.Sc. *Price 1s. 6d. net.*
No. 26.—On the Extent and Effects of Variety in Repetitive Work, by
H. M. Vernon, M.D., S. Wyatt, M.Sc., and A. D. Ogden,
M.R.San.I. *Price 1s. 6d. net.*
No. 27.—Results of Investigation in certain Industries. *Price 6d. net.*

B.—Metal and Engineering Industries.

- No. 1.—The Influence of Hours of Work and of Ventilation on Output
in Tinsplate Manufacture, by H. M. Vernon, M.D. *Price 6d. net.*
No. 2.—The Output of Women Workers in relation to Hours of Work
in Shell-making, by Ethel E. Osborne, M.Sc. *Price 6d. net.*
No. 3.—A Study of Improved Methods in an Iron Foundry, by C. S.
Myers, M.D., Sc.D., F.R.S. *Price 2d. net.*
No. 5.—Fatigue and Efficiency in the Iron and Steel Industry, by H. M.
Vernon, M.D. *Price 3s. net.*
No. 15.—Motion Study in Metal Polishing, by E. Farmer, M.A.
Price 2s. net.

C.—Textile Industries.

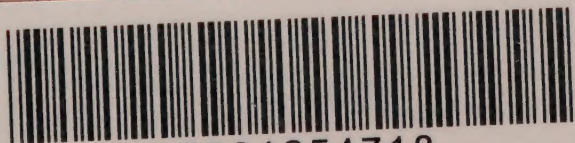
- No. 7.—Individual Differences in Output in the Cotton Industry, by
S. Wyatt, M.Sc., M.Ed. *Price 6d. net.*
No. 8.—Some Observations on Bobbin Winding, by S. Wyatt, M.Sc.,
and H. C. Weston. *Price 1s. 6d. net.*
No. 9.—A Study of Output in Silk Weaving during the Winter Months,
by P. M. Elton, M.Sc. *Price 2s. 6d. net.*
No. 17.—An Analysis of the Individual Differences in the Output of
Silk-Weavers, by P. M. Elton, M.Sc. *Price 1s. 6d. net.*

[CONTINUED ON PAGE 3 OF COVER.]

The above Reports can be purchased directly from H.M. STATIONERY
OFFICE at the following addresses :—Adastral House, Kingsway, London,
W.C.2; 28, Abingdon Street, London, S.W.1; York Street, Manchester;
1, St. Andrew's Crescent, Cardiff; or 120, George Street, Edinburgh; or
through any Bookseller.

The offices of the Board are unable to supply them directly.

Applications by post to the above addresses should quote the description
in full of the publications wanted, and should be accompanied by the price
indicated in the list.



22501254718

G. B. and I
MEDICAL RESEARCH COUNCIL.

INDUSTRIAL
FATIGUE RESEARCH BOARD.

REPORT No. 28.

The Function of Statistical
Method in Scientific
Investigation.

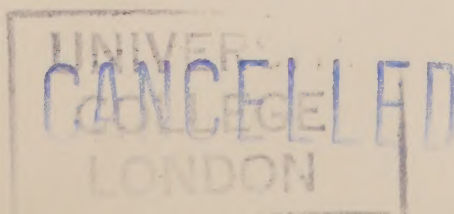
By G. UDNY YULE, C.B.E., M.A., F.R.S.

LONDON :

Published by His Majesty's Stationery Office, and to be purchased
at any of the addresses shown on opposite page.

1924.

Price 6d. Net.



THE BOARD.

WILLIAM GRAHAM, LL.B., M.P.—*Chairman*.

R. R. BANNATYNE, C.B. (Assistant Secretary, Home Office).

C. J. BOND, C.M.G., F.R.C.S.

W. L. HICHENS (Chairman, Messrs. Cammell, Laird & Co., Ltd.).

C. S. MYERS, C.B.E., M.D., F.R.S. (Director of the National Institute of Industrial Psychology).

Sir JOSEPH PETAVEL, K.B.E., D.Sc., F.R.S. (Director of the National Physical Laboratory).

Sir CHARLES SHERRINGTON, O.M., G.B.E., Sc.D., Pres.R.S. (Professor of Physiology, University of Oxford).

E. H. STARLING, C.M.G., F.R.S. (Foulerton Research Professor, Royal Society).

MONA WILSON, J.P.

D. R. WILSON, M.A., *Secretary*.

WELLCOME INSTITUTE LIBRARY	
Coll.	welMOMec
Call	pam
No.	WA400
	1924
	Y95f

TERMS OF REFERENCE.

To suggest problems for investigation, and to advise upon or carry out schemes of research referred to them from time to time by the Medical Research Council, undertaken to promote better knowledge of the relations of hours of labour and of other conditions of employment, including methods of work, to functions of the human body, having regard both to the preservation of health among the workers and to industrial efficiency; and to take steps to secure the co-operation of industries in the fullest practical application of the results of this research work to the needs of industry.

OFFICES :

15, York Buildings, Adelphi, London, W.C.2.

PREFACE.

The scientific investigation of many, and perhaps of most, industrial problems is necessarily founded largely on the application of statistical method, whether this consists in the simplest commonsense interpretation of numerical data or involves the employment of procedures of a highly technical character, the details of which are obscure and even repellant to those who are not specialists. A principal reason is that the investigators are only exceptionally able to rely upon the method of pure experiment. It is obviously impossible to modify at will the conditions under which industrial work is carried out; a factory is not an academic laboratory wherein the scientific investigator can eliminate from his experiment variables which he does not, at the moment, desire to study. Even when a measurable change is made in the conditions in a factory, the effect in any given respect of this change is nearly always obscured by the play of other factors not directly relevant—or not known to be directly relevant—to the matter under study. It follows that the student of such problems must take note of many things which the laboratory worker is able to eliminate, and is therefore obliged to rely more than the latter upon statistical methods.

Since in many of the researches published by the Board the statistical method of investigation is largely employed, the Board have long felt that the broad principles of this method and the reasons why it must be employed in their investigations should be more widely known, and that better knowledge of them would render their Reports of greater general value.

Such an explanation can only be furnished by one who, in addition to an expert knowledge of the subject, has the rare gift of being able to elucidate principles without recourse to technical detail. The qualifications of Mr. G. Udny Yule in both respects are well known, and the Board are very grateful to him for permission to include in their series of reports the following paper.

The Board think that a perusal of Mr. Yule's paper will satisfy the reader of the necessity of using statistical methods in the study of industrial fatigue, as well as of many other industrial problems, and will also make plain both the powers and the limitations of statistical methods of reasoning. In particular, it will be manifest that the caution displayed on so many occasions in drawing conclusions from what might appear to be very ample data is imposed by the complexity of the conditions,

and that in all investigations of a statistical character the validity of inferences must be tested in a manner which, to anyone not thoroughly familiar with the inherent difficulties of the subject, may seem over-cautious. Full comprehension of the principles expounded by Mr. Yule will save both investigators and the industrial world many disappointments, the inevitable retribution of hasty reasoning or of the neglect of relevant factors only rendered evident by careful study. It need not be supposed, however, that because the evidence contained in a particular report is inconclusive that the work has been thrown away. The cumulative effect of a series of testimonies, each indecisive by itself, may be very powerful. The accumulation of such necessarily incomplete evidence and its evaluation and integration by statistical methods are among the Board's most important duties.

October, 1924.

INDUSTRIAL FATIGUE RESEARCH BOARD.
REPORT No. 28.

**THE FUNCTION OF STATISTICAL METHOD IN SCIENTIFIC
INVESTIGATION.***

By G. Udny Yule, C.B.E., M.A., F.R.S.

The words "statistics" and "statistical" have undergone such marked changes since their introduction into the English language little more than a century and a half ago, that misunderstandings as to the function of statistical method in scientific investigation are natural, indeed almost inevitable.

Every newspaper reader knows, in a general way at least, what is meant *e.g.* by the term "vital statistics"; he sees nothing odd in such an expression as "meteorological statistics," and may not be very surprised at a statement, say, that an innings at cricket or a cricketing season has been remarkable "from the statistical standpoint." The reader who is familiar with scientific work will also have met with such phrases as "statistical studies in biology" and "statistical studies of the stars." If pressed for the meaning of the word as applied to such diverse cases, the general reader will, I suppose, probably say that statistics are long series of numerical observations, that "statistical studies" mean studies of groups or aggregates, or something of that sort. In any case, he will be sure that "statistics," "statistical," imply that the observations are numerical, and will not consent to regard as statistical, I feel confident, an account of a country dealing purely descriptively with its geography, mode of government, products and industries, the character of its inhabitants, their religion, and so on.

From the standpoint of the present meaning of the words he is right; from the historical standpoint quite wrong. Such a verbal description was precisely what was originally meant by a "statistical account" or by "statistics": the word meant "state-istics"—those matters which interest the statist or statesman. Inevitably an account of such matters was not at the time of the first introduction of the word into a European

* The following was delivered as a lecture at the University of Leeds in June, 1923. The MS. has been slightly revised for publication in its present form, but the original intention will, I hope, excuse the fact that the text still bears obvious traces of being addressed to an audience rather than to a reader.

language (namely, the German language, in the middle of the 18th century) a numerical account, for there were very few figures to give. The earlier works in English in which the words "statistics" and "statistical" were used were either translations from the German, written by Germans, or offshoots of the German school, and the words were naturally used in the German sense, carrying no necessarily numerical signification. They have at present been traced back to a translation, by W. Hooper, M.D., of a work by Baron von Bielfeld under the rather overwhelming title, "Elements of Universal Erudition" (1770). Even when the Royal Statistical Society was founded in 1834, the official definition had no reference to numerical methods. "Statistics," we read, "may be said, in the words of the prospectus of this Society, to be the ascertainment and bringing together of those facts which are calculated to illustrate the conditions and prospects of society"—a definition which might have been accepted by any member of the German school. It is, however, admitted that "the statist commonly prefers to employ figures and tabular exhibitions."

It was therefore only gradually and during the first half of the 19th century that the word "statistics" ceased to signify the science or art—I do not wish to quarrel as to the term—of describing the important characteristics of a state by means of the written word, and came chiefly to imply the art of describing such characteristics by means of numerical data. Further, as most of us will insist on drawing or attempting to draw conclusions from the data before us, there was a natural extension of the term to cover not merely the art of description, but the means of drawing such conclusions. In this sense, as indeed in its earlier sense, the word was on a par with such terms as "statics," "dynamics," or "mathematics," signifying a certain discipline, a sense in which it is rarely used now except in the phrase "theory of statistics." The transfer of the term from the discipline to the data themselves, to the modern sense, that is, in which we speak of "vital statistics" or "statistics of trade," appears to have taken place during the same period.

The development of the adjective "statistical" was naturally similar. The methods applied to the study of numerical data concerning the state were still called statistical methods, even when applied to data from quite other sources—what else should they be called when they were the same methods? Thus we now have a Journal (*Biometrika*) "for the statistical study of biological problems," and "statistical investigations" as to the behaviour of molecules or of the stars.

What is there, then, common to all these cases that the same methods should be called for? Very little consideration suggests the answer. The student of social facts cannot experiment, but must deal with circumstances as they occur entirely apart from his control. The numbers given by his "statistics" are pure observations, records simply of what has happened. The expert

in public health, for example, must take the records of deaths as they occur, and endeavour as best he can to interpret, say, the varying incidence of death on different districts. Clearly this is a very difficult matter. The proportion of deaths to population in a district is affected not only by its sanitary condition—however broadly we interpret that term—but by all kinds of other circumstances; not only by such definite circumstances as the ages of the population (if there are many of the old this will tend to throw up the number of deaths), and the proportion of the sexes (age for age, the mortality amongst women is usually less than amongst men), but also by that medley of circumstance which differentiates any chance sample of individuals from any other.

The purpose of *experiment* is to replace these highly complex tangles of causation, which confront the unfortunate investigator who is limited to pure observation, by quite simple systems in which only one causal circumstance is permitted to vary at a time. When this is done, the effects of changes in the one factor stand out clearly by themselves. When it cannot be done, the effect of changes in the one factor is overlaid by the effects of all kinds of other causes, “disturbing causes” as they may be called, for they are causes the operation of which we wish to exclude, and they disturb the simple effect of the one factor the influence of which we wish to note. Statistical methods are methods for handling and elucidating the meaning of data affected in this way by “disturbing causes,” or generally by a multiplicity of causes. Hence their very wide applicability.

The more perfect the experiment—the more nearly the experimental ideal is attained—the less is the influence of disturbing causes, and the less necessary the use of statistical methods. The more imperfect the experiment—the greater the failure to attain the experimental ideal—the greater is the need for statistical methods. Experiment is most perfect in the case of physics and chemistry. Here the influence of disturbing causes is usually small, though not negligible. In biology—in many instances at least—experiment is inevitably much less perfect. The external conditions are much more difficult adequately to control; the internal conditions of either plants or animals are largely beyond control, and beyond even such observation as may enable us, if we so desire, to select a series of similar individuals. Experimental work, therefore, on plants or animals—work, for example, in physiology, genetics, or psychology—is, as a rule, very far from attaining the experimental ideal: a residuum of disturbing causes is nearly always troublesome. In such branches of work as require the experiment—or let us call it the trial—to be carried out under practical conditions, and not under the more readily controllable conditions of the laboratory, the case is more difficult still. In agriculture, for example, it is hardly too much to say that experiment is nearly

always, from the standpoint of the ideal, abominably and outrageously bad and almost inevitably bad. Superposed on the difficulties of biological work in the laboratory are now the added difficulties due to the fact that work has to be carried out in the open field or the cattle shed. The greatest care is necessary even partially to eliminate the effects of disturbing causes, and the adoption of what may seem to be every possible precaution may prove disappointing, disturbing causes still exercising a large and almost a preponderating influence on the results.

But in agriculture experiment—the intentional alteration of conditions—is at least possible. In some lines of investigation not even this may be possible, or, if possible, may not be permitted. Investigations into the effect on the efficiency of labour of alterations in working conditions in factories are a case in point. Much important work on such questions is at present being conducted under the Industrial Fatigue Research Board, and it often presents very great difficulties. The investigator cannot play about with factory conditions, and alter them this way and that as he pleases in order to observe the effect of the alterations. All he can do is to take the records of Factories A, B, C, etc., which happen to have altered their hours of work, say, and see what he can extract from those records. The investigator is back in the position of the “statistician” in the older sense of the term—he is limited to simple observation, and there is precisely the same need that he should “think statistically” and use statistical methods with sense and skill.

Now before going further and enquiring rather more closely into the nature of statistical methods, does not this brief discussion lead to one or two rather useful morals? The experimenter who can attain to something near the experimental ideal has an easy job compared with the statistician. If he wants to know what effect an alteration in X has on Y , he simply alters X (without altering anything else) and looks. The unhappy statistician has to try to disentangle the effect from the ravelled skein with which he is presented. No easy matter this, very often; and a matter demanding not merely a knowledge of method, but all the best qualities that an investigator can possess—strong common sense, caution, reasoning power and imagination. And when he has come to his conclusion the statistician must not forget his caution: he should not be dogmatic. “You can prove anything by statistics” is a common gibe. Its contrary is more nearly true—you can never prove anything by statistics. The statistician is dealing with most complex cases of multiple causation. He may show that the facts are in accordance with this hypothesis or that. But it is quite another thing to show that all other possible hypotheses are excluded, and that the facts do not admit of any interpretation other than the particular one he may have in mind.

For another moral, let the experimenter who is driven to use statistical methods not forget this, that the very fact that he is compelled to use statistical methods is a reflection on his experimental work. It shows that he has failed to attain the very object of experiment and exclude disturbing causes. He should ask himself at every stage: Are these disturbing causes really inevitable? Can I in no way eliminate them or reduce their influence? This may be impossible: in such difficult experimental sciences as agriculture or psychology, the experimenter may have done the best that can be done without excluding more than a fraction of the effects of disturbing causes, and there is nothing definite to guide him as to when this "best possible" has been attained. But in any case it should always be the aim of the experimenter to reduce to a minimum the weight of statistical methods in his investigations. This may seem so obvious as hardly to require statement, but it is, I think, sometimes forgotten. Having acquired facility in the use of statistical methods, the investigator may be too apt, in the sheer joy of using a tool of which he is master, to neglect the adequate use of his best tool, which is, and will always remain, experiment. It is a fault which has not always been absent, as it seems to me, from work in psychology. It is often the case (*e.g.* in biology) that the experimental worker shows a certain, indeed a strong, prejudice against statistical work. From the present standpoint his attitude is natural and right, so long as perfect experiment or nearly perfect experiment is possible. But where this is no longer the case, the attitude ceases to be justifiable. Statistical methods afford the only hope of progress.

Now let us turn to a rather more detailed consideration of the sort of questions that theory of statistics can be called on to answer. Take a case, an agricultural case, in which multiplicity of causes is unavoidable. Suppose we examine a number of mangel roots and analyse them for the percentage of dry matter, which is an indication of their feeding value. For a particular sort the average works out at 14.6 per cent. on 160 roots analysed, but in different roots it ranges nearly all the way from 10 per cent. to 20 per cent. The graph (fig. 1) shows the "frequency distribution" as it is termed, *i.e.*, the number of roots with a percentage between 10 and 11, between 11 and 12, and so on. What, then, is the value of the average when we have got it? It is evident that, with such a range of variation, the analysis of a single root will be of little service; the average of two roots will be only little better; the average of, say, twenty (as large a number as will often be taken) better still. But how much better? Twenty roots is only a small sample out of the infinity of roots that might be analysed. How nearly may we expect the average of the twenty to approach the average of an indefinitely large sample? Now common sense is some guide as to the sort of thing we may expect; as already indicated, it suggests one rule: *an average is the more trustworthy, the greater the number of observations on which it is based.* But,

further, if the range of variation in the individual roots had been, not from 10 per cent. to 20 per cent., but only from 14 per cent. to 16 per cent., it is evident that we should have had a good deal more confidence in an average based on twenty roots. Common

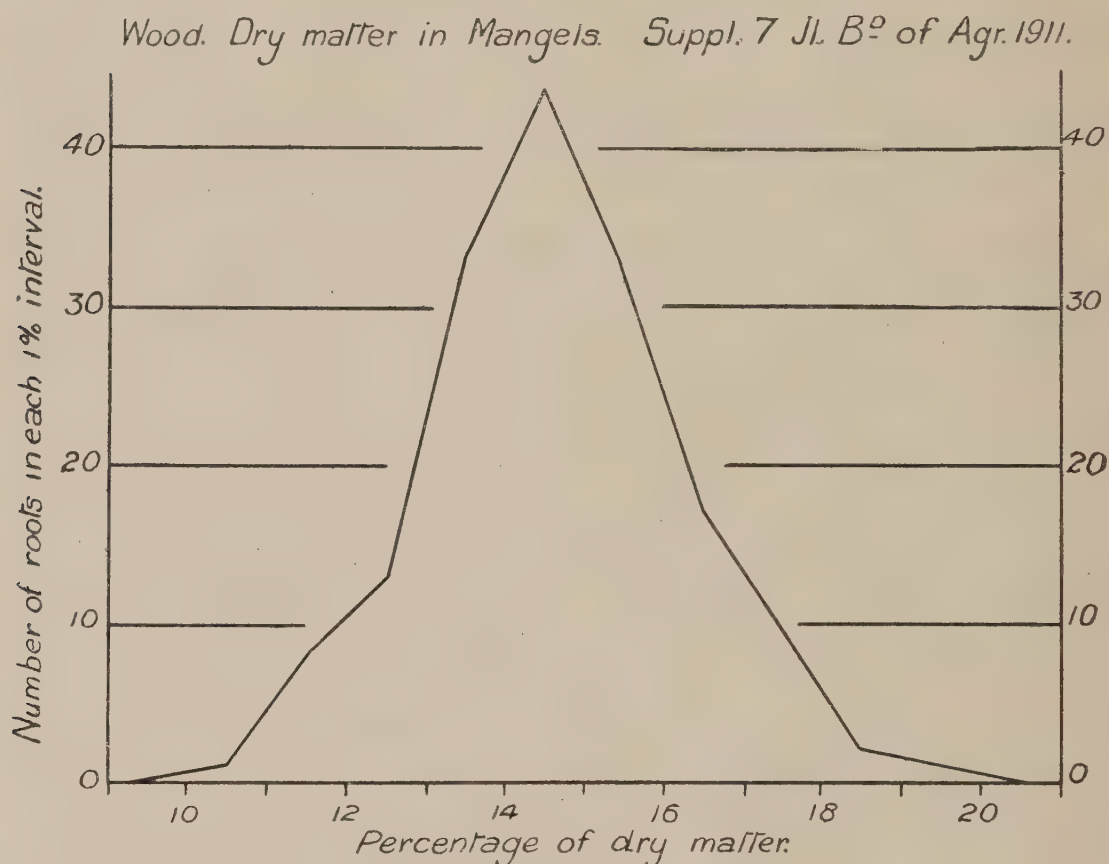


FIGURE 1.

sense, in fact, suggests the further rule: *an average based on a given number of observations is the more trustworthy, the less the original observations differ amongst themselves; the less trustworthy, the more they are scattered.*

The function of one important part of the theory of statistics—the theory of sampling—is to render such rules as these more precise: to show exactly in what way the trustworthiness of the average, for example, varies with the number of observations on which it is based and with the degree of scatter or *dispersion*, as it is called, of the original observations. In this way the investigator is assisted critically to estimate the value of his own results; he may be prevented from wasting his time by erecting some elaborate superstructure of argument on a difference between two averages which is no greater than a difference that might well be obtained on drawing two random samples from one and the same record; he can tell how many observations he must make in order to attain a given degree of precision in his average. But in order to do this, note that he must do something more than calculate an average; he must also calculate a measure of *dispersion*, and he must have a sufficient number of observations to give him that dispersion with reasonable precision.

Clearly the principle that I have endeavoured very briefly to indicate is a general one. It does not matter whether we are endeavouring to determine the average percentage of dry matter in a lot of mangel roots, the average yield of a particular sort of cereal, the average breaking load of a given length of a particular sort of yarn, the average time taken to perform some factory operation under given conditions, or the average observed expansion of a rod of metal for a given increment of temperature. It makes no difference that in any one of the former cases our average may be regarded by some as only an approximation to the average that would be obtained in an indefinitely large sample, and in the last case as an approximation to what we would more generally term "the true result." The principle is the same. We must repeat the observations, and we must calculate a measure of dispersion in order to form any idea of the *precision* of our result, of the limits within which we can trust it.

We must in such cases repeat the observations in order to determine not merely our average, but also our measure of dispersion, because we have no *a priori* knowledge of either. But in other cases we may have a theory as to the mechanism at work, and this theory may give us not merely the average and the measure of dispersion, but the entire "frequency distribution" of the variable; that is, it may tell us how often to expect values of the variable between the successive limits X , $X + h$, $X + 2h$, $X + 3h$, and so on; or it may happen that the theory can only tell us the general mathematical form of the frequency distribution, and not the particular arithmetical values that its mean and its dispersion will exhibit. In either case we have a test of the theory, very detailed in the first instance, less detailed in the second.

Thus the Mendelian theory of heredity tells us that, in a certain simple case of crossing (hybridising) an individual (plant or animal) possessing a characteristic A with another individual, a , not possessing that characteristic, we should expect in the second generation three A 's to one a , on an average; and that apart from this the fluctuations observed should be simply a matter of chance, like the fluctuations observed in drawing samples of n balls each from a large bucket containing a mixture of black and white balls in the proportion 3 to 1. The mechanism here is completely specified, and we can completely predict the average results that should follow if the theory is true; we can say, if, for example, we have data as to a large number of litters of 4 each, how often we ought to expect 4 A 's, 3 A 's, 2 A 's, 1 A , or no A 's at all in each litter. Sometimes a test of this kind shows an extraordinarily close consonance between theory and fact; sometimes there are odd divergences from expectation which raise further questions.

For an example of the second case, suppose we have a number of persons exposed to accident during a certain time, it may be months or it may be years; that each individual is equally likely

to meet with an accident during any element of time, and that we note how many of these individuals meet with 0, 1, 2, 3, . . . accidents during the time of exposure. Here, obviously, we cannot predict the actual average number of accidents per person to be expected, but we can predict the mathematical form of the distribution. During the war Dr. M. Greenwood, of the Ministry of Health, then in the Ministry of Munitions, was provided with certain data of the kind supposed as to girls in munition works; the "accidents" were trivial accidents, merely sufficient to send the girl to the welfare room, so that one girl might meet with even as many as six accidents or more during a few weeks' exposure to risk. Dr. Greenwood discussed the case with me,* and rather to our surprise, it was found that the frequency distributions did not follow the law expected. It was clear, then, that the girls could not be regarded as a homogeneous group all equally liable to accident; but the question remained, how *could* they be regarded in order to account for the observed form of frequency distribution?

The assumption was first tried that the girls could be regarded as forming two groups, a group more liable to accident and a group less liable; but still theory and fact were not satisfactorily in accordance. To assume a hypothetical division into three groups led to work that was almost too complex. It then suggested itself as possible that an accident so upset a girl that her chance of meeting with another accident was altered by the occurrence. The case was completely worked out, but the results were not rational, for it appeared that the girl's liability was first lowered and then raised by her meeting with accidents; and it must be admitted that the assumption was a very improbable one, since the accidents were quite trivial in character. We then returned to the first and more probable notion that the girls did not form a homogeneous group, but instead of endeavouring to break up the population into two, three, or more distinct groups with different liabilities, it was assumed simply that there was a certain frequency distribution of liability amongst the girls—that their liabilities to accident varied in the same sort of way that their statures or head-measurements or any other physical characteristics varied. A lucky shot at the possible form of this unknown frequency distribution led to a simple formula for the derived distribution of accidents, and very good agreement was now obtained between theory and fact.

It followed that, if the theory were right, and different girls had different liabilities, a girl who met with a large number of accidents during one period of exposure should tend to exhibit a large number of accidents during another, and conversely; the numbers of accidents met with by a girl during two successive periods of exposure being "positively correlated" as the statistician

* Greenwood and Yule, *J. Roy. Stat. Soc.*, March, 1920; and Greenwood and Woods, *Report No. 4 of Industrial Fatigue Research Board*, 1919.

terms it. The test was applied by Dr. Greenwood and the fact verified. A simple form of the test is to sort the girls into two groups, (a) those who meet with accidents during a trial period, (b) those who meet with no accidents during the same trial period ;

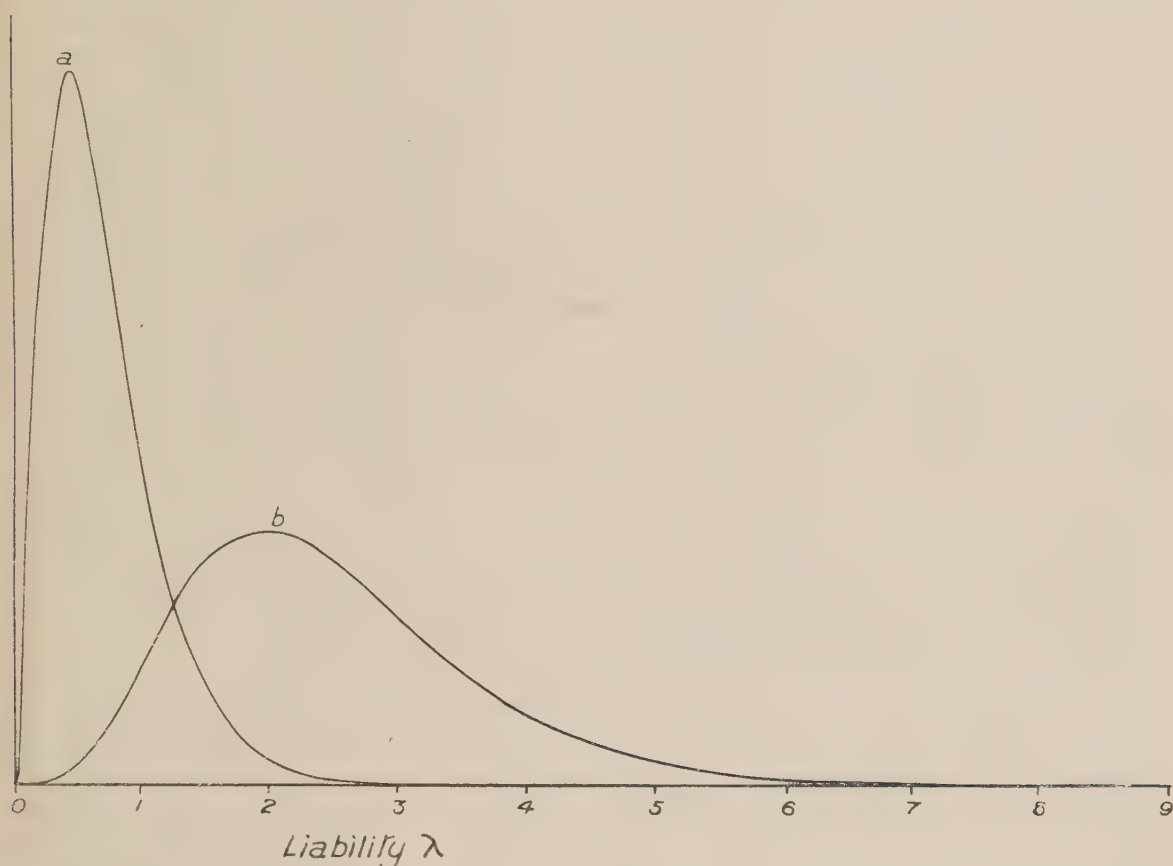


FIGURE 2.—Deduced distributions of liability λ (a) for the women who had not, and (b) those who had, accidents in January. (*Industrial Fatigue Research Board, Report No. 4, Table 9*).

and then form the frequency distributions for girls meeting with 0, 1, 2, 3, . . . accidents during a second period for each of the two groups separately. The result shows a striking difference, the girls meeting with no accidents during the trial period meeting with markedly fewer accidents during the second period of exposure to risk. This interesting piece of work has led to further investigations now being conducted under the Industrial Fatigue Research Board, to see whether it is possible by any form of psychological test to sort out the employees more liable to meet with accident. Fig. 2 shows the deduced frequency distributions of “liability to accident” for women who did not and who did meet with accidents respectively during the test month for the case of Table IX in Report 4 of the Board. The scale of “liability” is arbitrary, and the curves for the two groups are drawn so as to have the same area. It will be seen that the most frequent liability for the women who did not meet with accidents is between 0.4 and 0.5, and very few show a liability exceeding 2. For the women who *did* meet with accidents the most frequent liability is in the neighbourhood of 2, and some of them show liabilities exceeding 6. The two distributions differ so largely that a division by some form of test seems by no means hopeless.

On the assumption that the girls were a homogeneous group, each girl having the same liability to accident, it was expected that the frequency distribution would follow a certain mathematical law called the "Law of Small Numbers," or better the "Law of Small Chances." The expression for this law was first given by Poisson in 1837, in his *Recherches sur la probabilité des jugements*. It is an interesting illustration of the wide applicability of a statistical method in diverse fields that of recent years the law was twice worked out again by investigators ignorant of the work of Poisson, namely, in an appendix by H. Bateman to a paper by Rutherford and Geiger* on the emission of α -particles, and in a paper by "Student"† on the error of counting with a haemocytometer. The chance of a girl meeting with an accident during any element of time is small; the chance of the emission of an α -particle during any element of time is small; the chance of a particle falling into any assigned square on the haemocytometer is small; and as a consequence the same form of frequency distribution might be expected in all three cases. As yet another example, I may mention that the law found application by Dr. Greenwood and myself‡ to the interpretation of a certain test applied in the bacteriological examination of water.

When his work takes an investigator out of the field of nearly perfect experiment, in which the influence of disturbing causes is practically negligible, into the field of imperfect experiment (or *a fortiori* of pure observation) where the influence of disturbing causes is important, the first step necessary for him is to get out of the habit of thinking in terms of the single observation and to think in terms of the average. Some seem never to get beyond this stage. But the next stage is even more important, viz., to get out of the habit of thinking in terms of the average, and think in terms of the frequency distribution. Unless and until he does this, his conclusions will always be liable to fallacy. If someone states merely that the average of something is so-and-so, it should always be the first mental question of the reader: "This is all very well, but what is the frequency distribution likely to be? How much are the observations likely to be scattered round that average? And are they likely to be more scattered in the one direction than the other, or symmetrically round the average?" To raise questions of the kind is at least to enforce the limits of the reader's knowledge, and not only to render him more cautious in drawing conclusions, but possibly also to suggest the need for further work.

So far only those cases have been considered in which the problem related to a single attribute or a single variable, and the frequency of occurrence of different percentages of the attribute in samples of a given size, or of different values of the variable, is noted. But now suppose we note the occurrence of two or more attributes, or the values of two or more variables. All the pro-

* *Phil. Mag.*, 1910, 20.

† *Biometrika*, 1907, 5.

‡ *J. Hyg.*, 1917, 16.

blems of causal interpretation then arise, problems to which other sections of the theory of statistics—the theory of association, the theory of correlation, and so forth—are devoted. But these sections of the subject are more difficult to illustrate and to explain, and I must content myself with a very brief sketch.

The first case, where we simply note the presence or absence of some character is the simplest. We are making trial, say, of some new method of treating a disease—an inoculation, let us suppose. We simply note whether our patients were or were not inoculated, and whether they did or did not die. The simplest possible method of treatment of the data is here the best ; we compare the percentage of the uninoculated who survived with the percentage of the inoculated who survived. If the latter percentage is the larger, we *may* be able to conclude that the method of inoculation serves its intended purpose. But can we safely draw this conclusion ?

In the first place, since this is a new experiment, it is possible that the inoculation has only been tried on a dozen or two of cases. Now we know quite well that if we toss a coin some 24 times, we may sometimes get markedly fewer than the expected number of heads (12) and sometimes more—sometimes perhaps only 8 heads and sometimes 16. Hence if we have data for no more than some 24 cases inoculated and 24 uninoculated, it does *not* necessarily follow that even a fairly considerable difference between the percentages of deaths in the two groups really indicates any efficacy of the inoculation ; it may be merely a chance result, comparable to throwing first 8 heads and then 16 heads in two lots of 24 throws of a coin. The result must be controlled to see that it may not be (possibly at least) due to nothing more than the “chances of sampling.” Nor can we stop at this point. Even if the difference is so great that it lies clearly outside the limits of fluctuations of sampling, it still does not necessarily follow that it is due to the particular factor that we have so much in mind—the inoculation. If two attributes *A* and *B* are associated, this may be because *A* and *B* are both associated with some third attribute *C*. Is there any attribute *C* present in some cases and absent in others, which may possibly lead to fallacy ? We are making, perhaps, a rather uncertain experiment with an inoculation that threw some strain on the patient. Have we by any chance not liked to risk inoculation in the graver cases, and thus got an association between inoculation (*A*) and survival (*B*) merely owing to both being associated with lesser gravity of the case (*C*) ? At the risk of being blamed for damnable reiteration, let me repeat that the investigator *must* remember that, where he finds statistical methods to be necessary, many causes are at work and he must be cautious in his interpretations. Where some particular interpretation is rather attractive—it would no doubt be rather pleasant to him to believe that his inoculation is effective—he must be the more on his guard.

If the data give information, not merely as to the presence or absence of attributes, but as to the value of some variable X and some associated variable Y , more technical methods become possible. In the ideal experimental case, Y is some single-valued function of X ; we can plot a graph of the function and endeavour to determine its form, or we can find whether it is, in fact, some form of function suggested by theory. Where many disturbing causes are at work, Y is no longer a single-valued function of X . If we plot a point on squared paper for every pair of observed values of X and Y , we shall no longer get a series of points through all of which it is easy to run a smooth curve: we shall get a cloud of points, more or less widely scattered. If the dispersion is relatively slight, some form of single-valued functional relation may still be suggested, but if the dispersion is very wide *two* functional relations become important: the function expressing the average value of Y for a given value of X , and the function expressing the average value of X for a given value of Y . Where these two relations are both approximately linear we have the simplest case, and a case relatively more frequent than might be supposed. These approximate relations can be determined and a very useful coefficient can be calculated called the "coefficient of correlation," which can only range between $+1$ and -1 , its approach towards either limiting value indicating that the relation between X and Y approaches a simple linear law. The method of correlation can be extended to cover the case of several variables, an extension which is essential to the treatment of many problems where one variable is dependent simultaneously on two or more others.

As one illustration may be cited the exceedingly interesting and valuable work of Mr. R. H. Hooker* on the weather and the crops in England. Here the yield of the crop is dependent on at least two weather elements at any given period preceding harvest, the rainfall and the temperature, and Mr. Hooker has used the method of multiple correlation to analyse out the effect of each of these elements on all the principal crops of the Eastern Counties at different periods of the year. Mr. Hooker was, I think, the first to apply the method to this subject, but many others have followed both in this and in other countries. For another example from meteorology may be mentioned the work of Mr. C. E. P. Brooks† on the relation between the mean temperature at any point of the world, the percentage of land in a 10-degree circle to the west of the point, the percentage of land in the same circle to the east of the point, and the percentage of the circle covered by ice—four variables in all. As a third illustration, from the work of the Industrial Fatigue Research Board may be cited an investigation by Mr. S. Wyatt‡ into the relation between output, temperature, and relative humidity in a cotton weaving shed.

* *J. Roy. Stat. Soc.*, 1907, **70**; and *J. Roy. Met. Soc.*, 1922, **46**.

† *J. Roy. Met. Soc.*, 1918, **44**.

‡ *Report No. 23, Industrial Fatigue Research Board*, 1923.

The results of the investigation were suggestive rather than conclusive ; but this was not the fault of the investigator. It was probably due in part to the small range of conditions in the sheds tested, and in part to the large number of variables concerned.

The mere mechanical use of any such method is, of course, of little service ; when correlations and so forth have been calculated the real work of interpretation only begins, and the statistician must be prepared freely to adapt his methods to his problems. Let me take one example which particularly interested me a good many years ago.* The curve for the marriage-rate shows a series of oscillations or waves which rather closely reflect the general cyclical movement in trade and industry : it is required to investigate this relation more closely. As we are only concerned with the waves, and not with the long-period movements in the marriage-rate and trade, we must first of all isolate the waves from the remainder of the movement. Mr. Hooker suggested a very simple method for doing this† : in the case of the marriage-rate, for example, we may take the difference between the marriage-rate for each year and the mean rate for the nine or eleven years of which the given year is the centre. In the nine-year or eleven-year means the wave-movement is practically eliminated, and the differences consequently show up the waves apart from the slow movements. The same process can be applied to the trade-curve. Fig. 3 shows a graph of the results. Having isolated the waves, we can now correlate the ordinate of the marriage-rate wave

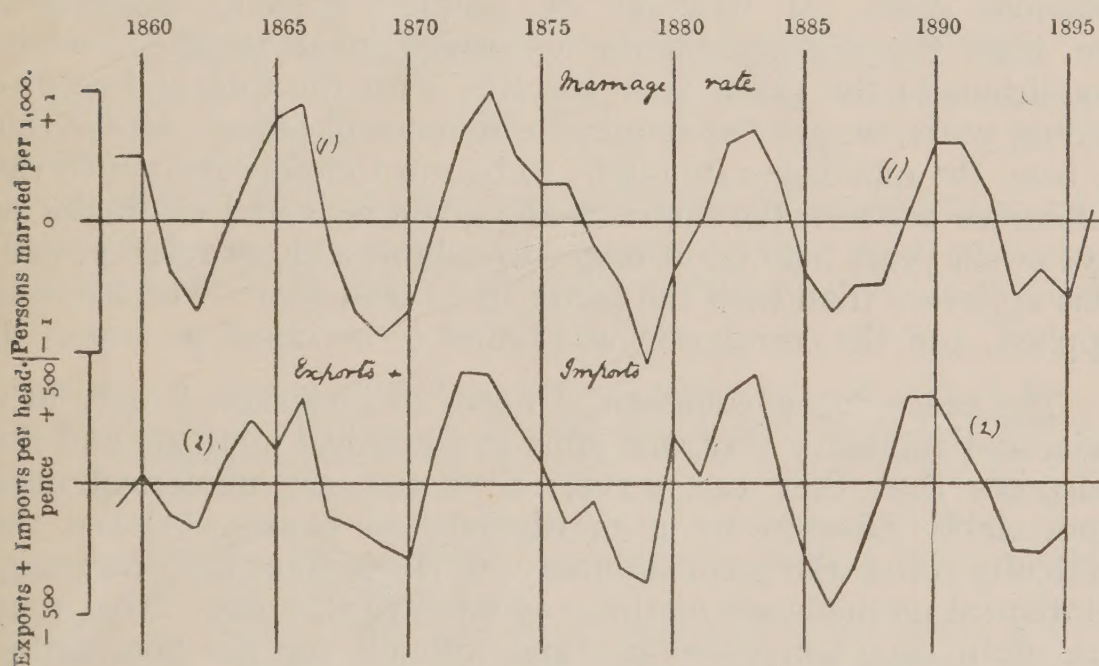


FIGURE 3.—Fluctuations in (1) Marriage-rate and (2) Trade (Exports + Imports per head) (Deviations from 9 year means). [Data of R. H. Hooker, *Journal of the Royal Statistical Society*, 1901].

not merely with the ordinate of the trade-wave in the same year, but with the ordinate of the trade-wave in the year before, or two years before, or a year after or two years after. In this way we can find for what difference of phase between the two waves the

* *J. Roy. Stat. Soc.*, 1906, 69.

† *Ibid.*, 1901, 64.

correlation is a maximum. As is suggested by mere inspection of the rather striking curves obtained in this way, it is found that the difference of phase is small, the marriage-rate waves only lagging slightly (a few months) behind the waves in the trade-curve.

At first sight the case seems quite simple. But consider : the people who do not marry this year because trade is bad have only postponed the happy event, and most of them will survive to marry when things take a turn for the better. The divergence of the marriage-rate from the normal will then depend on the *difference* between the postponements this year and the postponements the year before. If, then, postponements are regulated solely by this year's conditions, the marriage-rate should reach a maximum, not when the favourable factors are a maximum, but when they are increasing most rapidly, and that is some two to two and a half years before the maximum. The marriage-rate and trade waves ought to differ in phase by some two to two and a half years instead of by a few months only, and the marriage-rate wave ought to be in *advance* of the trade-wave and not lag behind it. This is not the fact. What is happening ?

The simplest solution of the difficulty that occurred to me is this, and it seems to me a fairly probable solution. The postponements in any one year do not depend on the trade conditions of that year only but on the conditions of that year and of several previous years. If, to make the simplest possible assumption, we take the postponements as proportional to the average conditions in the given year together with the four or five preceding years, we get the required difference of phase. But if this is true, the marriage-rate ought to be most closely related to the difference between the factor of the given year and the factor of five or six years before ; it ought to exhibit a closer relation with this *difference* than with the factor of its own year. The test was applied, and the correlation was found to be raised as expected.

The preceding is sufficient, I hope, to illustrate, if only in a brief and summary way, the aims of statistical methods and the purposes that they can serve. Statistics are numerical data appreciably affected by a multiplicity of causes. Hence the difficulty, often the great difficulty, of elucidating their meaning. Statistical methods are methods adapted to that end. That they are often complex, elaborate, and difficult for the non-mathematician to follow is unfortunate, but is an almost inevitable consequence of the complexity of the cases with which they endeavour, more or less imperfectly, to deal.

OTHER REPORTS OF THE INDUSTRIAL FATIGUE RESEARCH
BOARD—*continued.*

C.—*Textile Industries*—continued.

- No. 20.—A Study of Efficiency in Fine Linen Weaving, by H. C. Weston,
M.J.Inst.E. *Price 1s. 6d. net.*
No. 21.—Atmospheric Conditions in Cotton Weaving, by S. Wyatt,
M.Sc. *Price 2s. net.*
No. 23.—Variations in Efficiency in Cotton Weaving, by S. Wyatt,
M.Sc. *Price 3s. net.*

D.—*Boot and Shoe Industry.*

- No. 10.—Preliminary Notes on the Boot and Shoe Industry, by
J. Loveday, B.A., and S. H. Munro. *Price 1s. 6d. net.*
No. 11.—Preliminary Notes on Atmospheric Conditions in Boot and Shoe
Factories, by W. D. Hambly, B.Sc., and T. Bedford. *Price 3s. net.*

E.—*Pottery Industry.*

- No. 18.—Two Investigations in Potters' Shops, by H. M. Vernon, M.D.,
and T. Bedford. *Price 2s. 6d. net.*

F.—*Laundry Industry.*

- No. 22.—Some Studies in the Laundry Trade, by May Smith, M.A.
Price 2s. 6d. net.

G.—*Glass Industry.*

- No. 24.—A Comparison of different Shift Systems in the Glass Trade, by
E. Farmer, M.A., R. C. Brooks, M.A., and E. G. Chambers,
B.A. *Price 1s. 6d. net.*

The above Reports can be purchased directly from H.M. STATIONERY
OFFICE at the following addresses :—Adastral House, Kingsway, London,
W.C.2; 28, Abingdon Street, London, S.W.1; York Street, Manchester;
1, St. Andrew's Crescent, Cardiff; or 120, George Street, Edinburgh; or
through any Bookseller.

The offices of the Board are unable to supply them directly.

Applications by post to the above addresses should quote the description
in full of the publications wanted, and should be accompanied by the price
indicated in the list.

PAPERS BASED ON RESULTS OBTAINED FOR THE BOARD OR
ITS RELATED COMMITTEES AND PUBLISHED IN SCIENTIFIC
JOURNALS.

- BEDFORD, T. (1922) : The Ideal Work Curve.—*Jr. Ind. Hyg.*, **4**, 6.
CATHCART, E. P., BEDALE, E. M., and MCCALLUM, G. (1923) : Studies in
Muscle Activity.—I. The Static Effect.—*Jr. Physiol.*, **57**, 3 & 4.
CRIPPS, L. D., GREENWOOD, M., and NEWBOLD, E. M. (1923) : A Biometric
Study of the Inter-relations of Vital Capacity, Stature Stem
Length and Weight in a Sample of Healthy Male Adults.—
Biometrika, **14**, 3 & 4.
FARMER, E. (1922) : Time and Motion Study.—*Jr. Ind. Hyg.*, **4**, 4.
FARMER, E. (1923) : Interpretation and Plotting of Output Curves.—
Brit. Jr. Psych., **13**, 3.
GAW, F. (1923) : The Use of Performance Tests and Mechanical Tests in
Vocational Guidance.—*Jr. Nat. Inst. Ind. Psych.*, **1**, 8.
GREENWOOD, M., & NEWBOLD, E. M. (1923) : On the Estimation of
Metabolism from Determinations of Carbon Dioxide Production
and on the Estimation of External Work from the Respiratory
Metabolism.—*Jr. Hyg.*, **21**, 4.
HAMBLY, W. D., & MCSWINEY, B. A. (1922) : The U-tube Manometer
with relation to Muscular Exercise.—*Jr. Physiol.*, **57**, 1.

[SEE OVERLEAF.]

PAPERS BASED ON RESULTS OBTAINED FOR THE BOARD OR
ITS RELATED COMMITTEES AND PUBLISHED IN SCIENTIFIC
JOURNALS—*continued.*

- HEWITT, E. M., & BEDALE, E. M. (1923) : A Study of the Comparative Physiological Costs of Different Methods of Weight carrying by Women.—*Ann. Rep. of the Chief Inspector of Factories for 1922.*
- HILL, A. U., & LUPTON, H. (1923) : Muscular Exercise, Lactic Acid and the Supply and Utilisation of Oxygen.—*Quart. Jr. Med.*, **16**, 62.
- HILL, L., VERNON, H. M., and HARGOOD-ASH, D. (1922) : The Kata-Thermometer as a Measure of Ventilation.—*Proc. Roy. Soc.*, B. **93**, 198.
- LUPTON, H. (1922) : Relation between External Work and Time Occupied in a Single Muscular Contraction.—*Jr. Physiol.*, **57**, 1 & 2.
- LUPTON, H. (1923) : Exercise and Oxygen.—*Med. School. Gaz.*, **2**, 3.
- LUPTON, H. (1923) : An Analysis of the Effects of Speed on the Mechanical Efficiency of Human Muscular Movement.—*Jr. Physiol.*, **57**, 6.
- MUSCIO, B. (1920) : Fluctuations in Mental Efficiency.—*Brit. Jr. Psych.*, **10**, 4.
- MUSCIO, B. (1921) : Is a Fatigue Test Possible ?—*Brit. Jr. Psych.*, **12**, 1.
- MUSCIO, B. (1921) : Feeling Tone in Industry.—*Brit. Jr. Psych.*, **12**, 2.
- MUSCIO, B. (1922) : Motor Capacity with Special Reference to Vocational Guidance.—*Brit. Jr. Psych.*, **13**, 2.
- MUSCIO, B., & SOWTON, S. M. (1923) : Vocational Tests and Typewriting.—*Br. Jr. Psych.*, **13**, 4.
- NEWBOLD, E. M. (1923) : Note on Dr. Burnside's Paper on Errors of Observation.—*Biometrika*, **15**, 3 & 4.
- PEMBREY, M. S., & OTHERS (1922) : Tests for Physical Fitness.—Part I.—*Guy's Hospital Reports*, Oct., 1921.—Part II.—*Ibid.*, Oct., 1922.
- RUSHER, E. A. (1922) : The Statistics of Industrial Morbidity in Great Britain.—*Journ. Roy. Stat. Soc.*, **85**, 1.
- SMITH, MAY (1923) : Practical Problems of Industrial Life.—*Ind. Welfare*, **5**, 58.
- SMITH, MAY (1923) : The Laundry Trade.—*Laundry Record*, **33**, 388.
- STURT, M. (1921) : A Comparison of Speed with Accuracy in the Learning Process.—*Brit. Jr. Psych.*, **12**, Part 3.
- VERNON, H. M. (1921) : The Influence of Fatigue on Health and Longevity.—*Jr. Ind. Hyg.*, **3**, 3.
- VERNON, H. M. (1922) : The Influence of Rest Pauses and Changes of Posture on the Capacity for Muscular Work.—*Jr. Physiol.*, **56**, 47.
- VERNON, H. M. (1922) : Recent Investigation on Atmospheric Conditions in Industry.—*Jr. Ind. Hyg.*, **4**, 8.
- VERNON, H. M. (1923) : A Note on the Causes of Output Limitation.—*Jr. Nat. Inst. Ind. Psych.*, **1**, 5.
- VERNON, H. M. (1923) : Atmospheric Conditions in Potters' Shops and the Efficiency of Various Types of Drying Stoves.—*Trans. Ceramic Soc.*, **22**, p. 70.
- VERNON, H. M. (1923) : The Causation of Industrial Accidents.—*Jr. Ind. Hyg.*, **5**, 1.
- WILSON, D. R. (1921) : Some Effects of Environment on Fatigue and Efficiency.—*Jr. Roy. San. Inst.*, **42**, 3.
- WILSON, D. R. (1921) : The Work of the Industrial Fatigue Research Board and its Applications to Industry.—*Jr. Roy. Soc. Arts*, **70**, 3660.
- WILSON, D. R. (1923) : On Some Recent Contributions to the Study of Industrial Fatigue.—*Jr. Roy. Stat. Soc.*, **86**, 4.
- WILSON, D. R. (1923) : International Co-operation in the Study of Industrial Fatigue.—*International Labour Office, Studies and Reports, Ser. F.*, No. 9.
- WYATT, S., and WESTON, H. C. (1920) : A Fatigue Test under Industrial Conditions.—*Brit. Jr. Psych.*, **10**, Part 4.